505
961

Psychological Monographs

General and Applied

Connotative Meaning as a Determinant of Stimulus Generalization

By

Charles F. Dicken

University of Minnesota

Dicken

Price $1.50

l. 75
o. 1

Edited by Norman L. Munn

Published by the American Psychological Association, Inc.

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

NORMAN L. MUNN, Editor

Department of Psychology, Bowdoin College
Brunswick, Maine

Consulting Editors

ANNE ANASTASI
FRANK A. BEACH
ARNOLD M. BINDER
W. J. BROGDEN
ROBERT R. BUSH
JOHN F. DASHIELL
JAMES J. GIBSON
D. O. HEBB
EDNA HEIDBREDER
JOHN E. HORROCKS
FRANCIS W. IRWIN
JAMES J. JENKINS
BOYD MCCANDLESS
DONALD W. MACKINNON
QUINN MCNEMAR
LORRIN A. RIGGS
CARL R. ROGERS
DAVID SHAKOW
RICHARD L. SOLOMON
ROSS STAGNER

Manuscripts and correspondence on editorial matters should be sent to the Editor. *Psychological Monographs* publishes comprehensive experimental investigations and other psychological studies which do not lend themselves to adequate presentation as journal articles. Major space is given to the author's original contribution; introductory and bibliographic materials, as well as statistical tables and graphs, must be kept within reasonable bounds. Tables, graphs, and appendix materials which deal with detail not essential to adequate presentation of the findings may be made available through the American Documentation Institute—for details of this procedure, see the APA *Publication Manual.* Preparation of manuscripts for publication as monographs should follow the procedure given in the APA *Publication Manual.* Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

COPYRIGHT, 1961, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Vol. 75, No. 1 Whole No. 505, 1961

Psychological Monographs: General and Applied

CONNOTATIVE MEANING AS A DETERMINANT OF STIMULUS GENERALIZATION[1]

CHARLES F. DICKEN[2]

University of Minnesota

THIS study tests some implications of Charles Osgood's mediation theory of meaning (Osgood, 1952; Osgood, Suci, & Tannenbaum, 1957) and assesses the Semantic Differential (Osgood & Suci, 1955) as an instrument for the measurement of the meaning of verbal stimuli. The principal experiments concern mediated generalization among words selected for meaning relationships by the Differential.

Osgood (1952) defines "meaning" as a *representational mediation process*, a theoretical construct with the functional properties of an implicit, cue-producing response:

1. *Stimulus objects elicit a complex pattern of reactions from the organism, these reactions varying in their dependence upon the presence of the stimulus-object for their occurrence. . . .*

2. *When stimuli other than the stimulus-object (e.g., a sign) but previously associated with it are later presented without its support, they tend to elicit some reduced portion of the total behavior elicited by the stimulus object. . . .*

3. *The fraction of the total object-elicited behavior which finally constitutes the stable mediation process elicted by a sign will tend toward a minimum set by the discrimination capacity of the organism.* This is because the sole function of such mediating reactions in behavior is to provide a distinctive pattern of self-stimulation (cf. Hull's conception of the "pure stimulus act").

4. *The self-stimulation produced by sign-elicited mediation processes becomes conditioned in varying strengths to the initial responses in hierarchies of instrumental skill sequences.* This mediated self-stimulation is assumed to provide the "way of perceiving" signs or their "meaning" as well as mediating instrumental skill sequences—behaviors to signs which take account of the objects represented (pp. 203-204).

Figure 1a illustrates the conditioning of a sign ($\boxed{S}$) to the mediation process ($r_m s_m$) when $\boxed{S}$ occurs continuously with R_T, the total response of the organism to a stimulus-object Ṡ. Figure 1b shows the conditioning of the sign and the mediator to an instrumental response R_X.

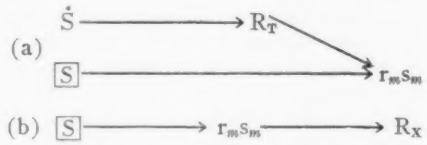


FIG. 1. Meaning as representational mediation process.

Two or more signs acquire similar mediation processes if they are associated with Ṡs which elicit similar R_T's. Such signs are considered to have similar meanings. Mediated stimulus generalization would be expected to occur among meaningfully similar signs because of the similarity of the cue properties of the mediators, even though the signs are not related in any primary stimulus dimension.

Figure 2 illustrates generalization between two signs with similar mediation processes. A mediation paradigm such as this can be used to explain the generalization observed among meaningfully similar words in conditioning experiments (Diven, 1937; Razran, 1935–36, 1949; Riess, 1940) and in studies of transfer and retroaction (Foley & Cofer, 1943; Haagen, 1943; McGeoch & McDonald, 1931; Morgan & Underwood, 1950; Osgood, 1946; Young & Underwood, 1954; Yum, 1931). Previous

[1] Drawn from a doctoral dissertation submitted to the University of Minnesota. The author is grateful to James Jenkins and Wallace Russell for advice and criticism. A report based on some of the data was read at the meeting of the Western Psychological Association in Monterrey, California, 1958.

[2] Now at the University of Chicago.

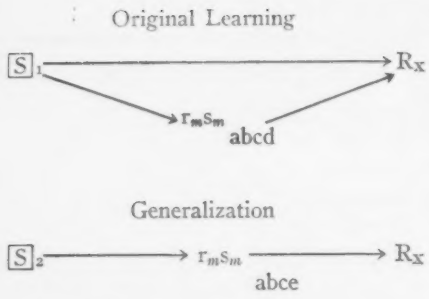


FIG. 2. Mediated stimulus generalization between signs of similar meaning.

experimentation has been limited by lack of a unifying conceptualization and by absence of satisfactory quantitative treatment of the meaning variable. The authors cited assessed meaning relationships by formal dictionary definitions, similarity of logical or content category, identity of meaning in different languages, or by judgments of similarity of meaning or feeling tone. The most refined technique (Haagen, 1943) employed scaled judgments of meaning similarity in sets of adjectives. No study used comparisons of *independent* measurements of the meaning properties of individual words in determining meaning relationships.

The Semantic Differential has been described in detail elsewhere (Osgood, 1952; Osgood & Suci, 1955; Osgood, Suci, & Tannenbaum, 1957). The underlying assumptions (Osgood, 1952) are stated as follows:

1. *The process of description or judgment can be conceived as the allocation of a concept to an experiential continuum, definable by a pair of polar terms. . . .*

2. *Many different experiential continua, or ways in which meanings vary, are essentially equivalent and hence may be represented by a single dimension. . . .*

3. *A limited number of such continua can be used to define a semantic space within which the meaning of any concept can be specified* (p. 227).

Judges typically rate words or other stimulus items on each of several seven-step continua whose end points are defined by "polar" adjective pairs. Osgood and Suci (1955) found that most of the covariance of 50 polar scales used in obtaining judgments of the meaning of 20 verbal concepts could be accounted for by a small number of factors. Three major factors which appeared in each of two analyses were labeled evaluation, potency, and activity. Osgood and Suci interpreted their findings as confirming Assumptions 2 and 3 above. They considered the meaning dimensions identified to be primarily *connotative* in character, and admitted that such dimensions may not assess other semantic attributes (e.g., denotation) which are probably much more numerous.

Appendix A shows a 20-scale form of the Differential constructed by Jenkins, Russell, and Suci (1958). Eight scales represent primarily an evaluative dimension, and each of the other major factors, potency and activity, are represented by three scales. The remaining scales relate partly to the major factors and partly to five minor factors. Factor loadings and communalities of the scales (Jenkins et al., 1958) are shown in Appendix B.

The set of judgmental responses to a word obtained by use of a multiscale form of the Differential determines a connotative meaning profile for the word. Let this set of responses be termed R_{SD}. When R_{SD} is a mean profile obtained from a group of judges, it represents the stable, consensual components of the word's connotative meaning within the linguistic-cultural group from which the judges were drawn. Similarity of meaning of words can be estimated by comparison of these profiles. Mean profiles for two connotatively similar words are plotted on the scales in Appendix A.

Osgood and Suci (1952) defined an index of profile similarity.

$$D = \sqrt{\sum_{i}^{n} d_{jl}^2}$$

where d_{jl} is the difference between two scaled stimulus items on Scale i, and n is the number of scales. The index reflects profile differences in absolute elevation or position as well as configural differences. In the present context, D may be used to specify the connotative dissimilarity of a pair of words: i.e., the words' semantic

"distance" in a 20-dimensional space defined by the scales used to determine the mean profiles. Small values of D reflect similarity of mean profile and hence similarity of measured connotative meaning.

Three related hypotheses are examined in the present study: (*a*) The meaning of a word stimulus may be identified with a representational process, ($r_m s_m$), which functions to mediate generalization among meaningfully related stimuli. (*b*) R_{SD} is an operational index of $r_m s_m$; words evoking similar R_{SD}'s evoke similar $r_m s_m$'s. (*c*) The Osgood-Suci D applied to Semantic Differential profiles is a measure of similarity of R_{SD} and hence of $r_m s_m$.

If these hypotheses are correct, words whose Semantic Differential profiles are similar (low values of D) would be expected to demonstrate mediated stimulus generalization in the manner illustrated in Figure 2. The mediator $r_m s_m$, the meaning measurement R_{SD}, and the similarity index D are central constructs in the Osgood meaning theory. Evidence of generalization among words selected solely for scaled similarity values of Semantic Differential profiles would substantiate the usefulness of these constructs.

Initial evidence on the relationship of R_{SD} to verbal generalization was reported by Ryan (1957). He compared the effect of previously established word association habits with that of similarity of Differential profile (measured by D) in facilitation of transfer in paired-associate learning. Interpolated list stimulus words which were related by association habits to original list stimuli demonstrated the greatest transfer. Transfer was, however, also observed for interpolated list stimulus words related to original list stimuli only by similarity of Differential profile. Ryan's data suggested the profitability of further investigation of R_{SD} and verbal generalization, and indicated the need for attention to the associative factor.

A technique which lends itself readily to problems of generalization among verbal stimuli had been worked out by Mink (1957). In a study of associative determinants of verbal generalization, Mink trained subjects to press a lever in response to words which were the stimulus terms (X) of word pairs which had been found associatively related in a free-association study (Russell & Jenkins, 1954). He found that the lever response generalized to the response terms (Y) of the ($X \rightarrow Y$) associative pairs. Mink's experimental procedures were adapted for the present experiments. Details of his findings will be referred to later in relation to the role of the associative factor in the present study.

Method

Six generalization experiments were performed with the same design and procedures. In an original learning phase, subjects performed a simple lever-press response upon visual presentation of each of a set of word stimuli. Trials were rapid and few to preclude complete learning. The original words, words similar in meaning to these, and control words were then presented as a generalization test. Subjects were instructed to respond to words they believed they recognized from the original set. Response to words similar in meaning to the original words in excess of response to control words was the measure of meaning-determined stimulus generalization. Details of procedure common to all experiments follow; variations introduced in individual experiments and procedures used in supplementary studies of the stimuli are discussed later.

Selection and Arrangement of Stimuli

The stimuli were drawn from a population of 360 verbal concepts[3] for which Jenkins, Russell, and Suci (1958) had obtained mean Semantic Differential profiles. Each mean profile was determined by the judgments of 15 male and 15 female college students recorded on the scales shown in Appendix A. Test-retest reliability of mean scale values for 20 randomly selected concepts was .97.

Jenkins, Russell, and Suci (1959) compared all possible pairs of concepts in the 360-word population (64,620 pairs) with the Osgood-Suci D index.

[3] The majority of the concepts were single words. The small number of two-word concepts included in the population was disregarded in selecting stimuli for the present experiments.

TABLE 1

WORDS AND INTRACLUSTER D VALUES

(Clusters A, B, M, and N)

Rank	Word	1	2	3	4	5	6	7	8
				Cluster A					
1	WOMAN								
2	FRAGRANT	2371							
3	PRETTY	2237	2420						
4	GIRL	1868	2424	1940					
5	LOVELY	2530	2013	2248	2798				
6	CHARMING	2092	2363	2241	2790	2229			
7	BODICE	2114	2323	2428	2028	2848	2951		
8	SONG	2644	2127	2683	2770	2673	2480	2458	
	$\overline{D}$	2265	2292	2313	2374	2444	2449	2450	2543
	Cluster mean	2391							
				Cluster B					
1	TROUBLE								
2	HURT	1612							
3	SCORCHING	2233	2473						
4	PAIN	1998	1829	2893					
5	MAD	2327	2561	2631	2904				
6	DANGER	2573	2466	2634	2448	2704			
7	SCALDING	2891	2764	2045	3087	2658	2525		
8	FRIGHTFUL	2732	2830	2597	2772	2200	3209	2651	
	$\overline{D}$	2338	2362	2501	2562	2569	2651	2660	2713
	Cluster mean	2543							
				Cluster M					
1	PROGRESS								
2	EFFORT	1572							
3	STUDY	2223	1681						
4	INCOME	2162	1957	2533					
5	MONEY	2563	2176	3082	2264				
6	UP	3009	2168	2401	3061	2856			
7	PATRIOT	2490	2282	2944	3007	3047	2842		
8	LIFT	3037	2405	3067	2931	2363	2242	2619	
	$\overline{D}$	2249	2381	2562	2563	2617	2654	2747	2752
	Cluster mean	2566							
				Cluster N					
1	BAD								
2	THIEF	2267							
3	CROOKED	2143	2556						
4	FRAUD	2089	2334	2110					
5	STEAL	1904	1680	2719	2496				
6	CRIMINAL	2805	1630	2787	2855	2641			
7	NASTY	2517	3052	2423	2701	3181	2724		
8	HATE	2325	2915	3120	3069	3311	2496	2531	
	$\overline{D}$	2331	2382	2471	2501	2591	2647	2687	2824
	Cluster mean	2554							

Note.—D values are shown without the decimal point which would follow the first digit.

Values of the index for pairs drawn from the concept population vary from less than 1.3 for highly similar items to more than 12.0 for highly dissimilar words. Norman (1959) found that D values for comparisons of mean profiles for the *same* concept obtained from two samples of judges varied from 1.28 to 2.92. He interpreted the data as suggesting that D values less than 2.00 should not be regarded as evidence that two concepts differ in measured connotative meaning. Norman also found that between-concept D values correlated .92 in two samples of judges, indicating a high degree of stability for relative values of D obtained in comparisons of mean profiles.

Four clusters of eight words with uniformly high intracluster connotative similarity were derived from the 360-word population by an arbitrary iterative procedure.[4] Table 1 shows the clusters, the intracluster D values for each pair, the average intracluster D for each word and the rank of this average, and the mean intracluster D.

Table 2 shows the elements used in the learning and test phases of a typical experiment. Two eight-word clusters, termed "experimental words" and designated, for example, A and B were used in each experiment. The prime and double prime (e.g., A', A'') indicate subclusters formed by assigning words ranking 1, 3, 5, and 7 in each eight-word cluster to the learning list and the remainder to the test list. The two test list subclusters, eight words in all, served as generalization stimuli and are termed G words. Meaning-determined stimulus generalization was expected to occur between each learning list subcluster and its test list counterpart in a manner analogous to that shown in Figure 2, except that the conditioning of an instrumental response to the mediator would presumably be quadruply reinforced by the four related L list words, and manifested upon presentation of each of the four corresponding G words.

The eight control (C) words provide a baseline for determining response to G words in excess of that expected on the basis of chance or factors other than the hypothesized generalization. The C words were selected for two properties: (*a*) Absence of connotative similarity in relation to all 16 experimental words. Absence of similarity was defined by $D \geqq 6.00$[5] for each C word in relation to every experimental word. (*b*) Comparability with the eight G words in Thorndike-Lorge (1944) usage frequency.

TABLE 2

COMPONENTS OF TYPICAL EXPERIMENT

Learning (L) list		Test (T) list		
Cluster A' words	4	Cluster A'' words	4	G words
Cluster B' words	4	Cluster B'' words	4	G words
Filler (F) words	4	Control (C) words	8	
	12	L list words	12	
			28	

Filler (F) words were used to expand the learning list to reasonable length and to reduce the chance of conscious recognition of the meaningfully related clusters. The F words were selected for $D \geqq 6.00$ in relation to all C words, to reduce the possibility of augmentation of response to the latter by generalization from the F words.

Each L list was arranged in three sequences so that each item occurred once in each third and no more than two words of the same class (A', B', or F, for example) occurred in sequence. Six different random orders of the T list were prepared.

Subjects

Twenty male undergraduate psychology students were subjects in each experiment. The subjects received two points final examination credit for participating.

Apparatus and Procedure

Each subject was tested individually. The subject was told the experiment concerned memory for words, recognition time, and reaction speed. One order of the L list was read aloud, one word per second, and then the other two orders were presented visually at the same rate on a

[4] The procedure consisted of successive application of a set of rules devised to locate words which, when added to an initial cluster of size two, would yield clusters of the desired size with a minimum intracluster D. Consideration was given to obtaining two clusters each from opposite ends of the major factorial dimension, evaluation. The experimenter applied the rules to tabular data furnished by Jenkins, Russell, and Suci (1959) in which words were identified by number only, eliminating the possibility of a subjective factor in selection of "similar" words and insuring equal likelihood of inclusion of any word with the desired quantitative properties.

[5] Since stimulus generalization has been demonstrated for antonyms (Cofer, Janis, & Rowell, 1943), very large semantic distances of C words in relation to L list words were avoided insofar as possible. Any failure to avoid the antonym-based stimulus generalization to C words is of course a conservative error with respect to the use of response to C words as a baseline for determining generalized response to G words.

standard Hull-type memory drum. During the visual presentation, the subject grasped a lever movable in a 60° arc and was instructed to press the lever through the full arc upon viewing each word. Before both auditory and visual presentations, the subject was told to try to remember the words so as to be able to respond to them later.

The six orders of the T list were next presented visually at the rate of one word per second. The subject was told to press the lever as quickly as possible in response to each learning list word he recognized or believed he recognized. The movements of the lever were relayed to a Sanborn recorder set for sensitivity to even slight pressures. Response amplitude was recorded linearly. Preliminary experimentation indicated no consistent amplitude differences for any of the stimulus classes, and the records were subsequently scored on the basis of presence or absence of response of any magnitude to each T list word.

Statistical Treatment of Response Data

The total number of responses of the 20 subjects to each T list word was computed and divided by the maximum total of 120 (six repetitions of each word for 20 subjects) to give a percentage response value for each word. In addition, four measures of generalization were computed for each subject: (*a*) total generalization, G_T, determined by subtracting total C word response from total G word response; (*b* and *c*) subcluster generalization, computed by subtracting total C word response from twice the total level of response to a four-word subcluster (e.g., $G_A = 2A'' - C$); and (*d*) differential subcluster response D_B (e.g., total A'' response minus total B'' response). Means and variances of each of these measures were computed, and the statistical significance of the deviation of each mean from zero was evaluated by a two-tailed t test with 19 *df*. A test of the significance of response to individual G words is described later.

Two Generalization Experiments

Experiment 1

Clusters A and B were used in the first experiment in the manner illustrated in Table 2. Table 3 shows the words used in the experiment and their assignment to the L and T lists. Percentage of possible response for each word on the test trial is also shown. The percentage response by stimulus class for the crucial T list words was 29.6 for A'' words, 14.0 for B'' words, and 7.5 for C words. The average total generalization value ($\bar{G}_T$) was 6.58 ($p < .001$). The average generalization value for A'' words ($\bar{G}_A$) was 9.95 ($p < .001$). Average generalization for B'' words ($\bar{G}_B$) was 3.10 ($p < .01$) and the average differential generalization value for the two subclusters of G words ($\bar{D}_G$) was 3.42 ($p < .01$).

The hypothesis of stimulus generalization to T list words connotatively similar to the original words was confirmed by the data of the first experiment. Two equally evident findings had not been expected: the two subclusters of G words differ significantly in response frequency, and individual G words vary widely in response frequency. Nothing in the generalization hypothesis or in the procedures for selecting the stimuli indicated that one cluster should yield more generalization than another, or that more than random variation should occur in responses to individual G words.

TABLE 3

Percentage Response to Individual Words: Test Trial, Experiment 1

L list words	% response	Words unique to T list	% response
A'		*A''*	
BODICE	99	CHARMING	57
LOVELY	80	GIRL	32
PRETTY	78	FRAGRANT	25
WOMAN	78	SONG	04
		All A'' words	29.6
B'		*B''*	
SCALDING	96	FRIGHTFUL	28
SCORCHING	96	PAIN	12
MAD	76	DANGER	09
TROUBLE	50	HURT	07
		All B'' words	14.0
F		Control	
SUCCESS	96	FAT	11
CAR	72	PLAIN	11
FOOD	54	SLOW	10
AFRAID	47	SNAIL	07
		SQUARE	07
		TRUNK	07
		DIM	04
		LONG	02
		All C words	7.5

Examination of the available data for the individual stimuli failed to yield an explanation of these phenomena. Cluster *B* is slightly less connotatively similar internally than Cluster *A* (Table 1), but the difference is very small, and does not seem to plausibly account for the large difference in generalization to *A''* and *B''*. The F words are slightly more closely related connotatively to Cluster *A* than to Cluster *B* (the *D* means for the two 4 × 8 word comparisons are 7.15 and 8.10, respectively) but the difference is again small, and the absolute values of *D* are so large as to question reference to "similarity" of F words and experimental words. The average Thorndike-Lorge frequency of the *A''* words is slightly *less* than the average for the *B''* words. The small intracluster differences in average *D* values for individual words and the Thorndike-Lorge frequencies do not relate systematically to the differences in individual word response frequency.

Experiment 2

Experiment 2 was performed to replicate Experiment 1 for subjects and verbal materials, in order to determine the stability of the main generalization effect and of the observed variability in response to subclusters and individual words. Clusters *M* and *N* were assigned as experimental words. A new set of control words and four new filler words balanced closely for similarity to the *M* and *N* clusters were selected. The components and response percentages for Experiment 2 are shown in Table 4.

Subcluster *N''* received 35.4% of the possible response, Subcluster *M''* 19.2%, and the C words 9.4%. $\bar{G}_T = 8.60$ ($p < .001$); $\bar{G}_M = 4.70$ ($p < .01$); $\bar{G}_N = 12.45$ ($p < .001$); $\bar{D}_G = 3.90$ ($p < .001$).

The main generalization hypothesis was again confirmed. The "cluster predominance" effect seen in the first experiment appeared again, and although the data were somewhat more encouraging as to uniformity of individual G word response, two members of *M''* fall far below the other G words in response frequency.

TABLE 4

PERCENTAGE RESPONSE TO INDIVIDUAL WORDS: TEST TRIAL, EXPERIMENT 2

L list words	% response	Words unique to T list	% response
M'		*M''*	
PATRIOT	98	EFFORT	38
PROGRESS	82	INCOME	21
MONEY	67	UP	10
STUDY	57	LIFT	07
		All *M''* words	19.2
N'		*N''*	
BAD	90	HATE	41
STEAL	71	THIEF	38
NASTY	71	CRIMINAL	32
CROOKED	60	FRAUD	30
		All *N''* words	35.4
F		Control	
ABRUPT	85	SLEEP	17
FEAR	82	STOUT	12
BLOCK	73	DUSKY	11
CITY	72	PIG	08
		ROUND	08
		CALM	08
		MOLD	06
		SOFT	05
		All C words	9.4

Response to Individual G Words

Inspection of the data from the first two experiments showed that G word response frequencies tended to fall in two "clumps" rather than in one highly variable distribution. The G words seemed to "work" or "not work" as generalization stimuli. A test of this possibility was carried out as follows. The distributions of frequency of response to the control words in the two experiments did not differ significantly in means or variances, and these distributions were combined to form a single distribution of frequencies of response to control words with $N = 16$. The mean and standard deviation of this pooled distribution were taken as estimates of the parameters of a population of control word response frequencies. None of the six G words in Experiments 1 and 2 with the lowest response frequencies (SONG, HURT, LIFT, DANGER, UP, and PAIN)

have response frequencies which deviate more than one standard deviation from the mean of this pooled C word distribution. In contrast, all of the remaining 10 G words fall more than 3.5 standard deviations above the mean of the pooled C word distribution, a deviation expected less than 4 times in 10,000 in cases drawn from a population with parameters defined by the C word distribution. Thus, the first group of G words is essentially indistinguishable from the C words in response frequency, while the second group is clearly distinguishable from the C words in this respect. The first group of 6 G words may be considered "nonfunctional" with respect to generalization; the other 10 G words appear clearly "functional" in this respect.

Confirmation of the hypothesis that only some of the G words functioned as generalization stimuli directed attention to the possibility that associative habits not reflected in the connotative meaning measurement might account for some or all of the obtained generalization.

Associative Relationships of the Generalization Stimuli

Generalization between words related by previously established associative habits has been repeatedly demonstrated. Investigators have assumed pre-existing verbal associations of the form $(X \rightarrow Y)$ from the frequent occurrence of Y as a response to X in normative studies of free association (Kent & Rosanoff, 1910; Russell & Jenkins, 1954). Ryan (1957) found that training with a series of paired verbal items $X \rightarrow A$ facilitated learning of a second series $Y \rightarrow A$ when $X \rightarrow Y$ associations were frequent in the Minnesota (Russell & Jenkins, 1954) norms. Bastian (1957) found that high frequency $X \rightarrow Y$ associations inferred from the same norms facilitated learning of $A \rightarrow Y$ pairs when $A \rightarrow X$ pairs had been previously learned. Relearning of the $A \rightarrow X$ pairs following the interpolated task was also facilitated by $X \rightarrow Y$ associations. McClelland and Heath (1943) found interference in the relearning of $X \rightarrow A$ pairs when a $Y \rightarrow B$ task was interpolated and $X \rightarrow Y$ associations were frequent in the O'Connor (1928) norms. Mediated generalization of the form $X \rightarrow (Y) \rightarrow B$ was assumed to retard relearning of the correct $X \rightarrow A$ pairs.

Mink's (1957) extensive study of stimulus generalization between unidirectional verbal associates (X elicits Y associatively, but Y does not elicit X) is particularly pertinent. Generalization of lever-press responses from X words used in the learning list to Y words used in the test list was demonstrated in four experiments. The generalization occurred under varying rates of presentation, differing list arrangements, and different strengths of the association variable. A regular relationship between association strength and amount of generalization was observed for associates of medium and high frequencies, although no generalization occurred between $X \rightarrow Y$ pairs with less than 30% normative association frequency. In accounting for his findings, Mink assumed an implicit occurrence of a response like pronouncing the Y word at the time of the ($X \rightarrow$ lever press) learning trials. This was inferred on the basis of the tendency of Y to occur associatively to X. If this were so, some strength would presumably then accrue to habits of the form ($Y \rightarrow$ lever press), and result in response to Y when it appeared on the generalization trial. Figure 3 shows these relationships, which are termed the Mink paradigm. X and Y are unidirectional associates as described above, (Y) is the hypothesized implicit occurrence of the Y response, and R_1 is the lever-press response.

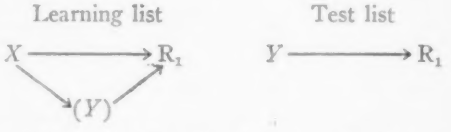


Fig. 3. Mink paradigm for associatively-determined generalization.

Mink also conducted four experiments in which the X and Y words were deployed in what he termed the Shipley-Lumsdaine mediation paradigm (Lumsdaine, 1939).

Figure 4 illustrates this arrangement and the expected generalization of the lever response from Y to X. Contrary to expectation, and in spite of several variations in procedure, there was no generalization to the X words. Mink concluded that direction as well as strength of association is crucial in generalization mediated by associative habits.

Learning list Test list

$Y \longrightarrow R_1 \quad X \longrightarrow (Y) \longrightarrow R_1$

FIG. 4. Shipley-Lumsdaine paradigm for associatively-determined generalization.

The clear evidence for stimulus generalization to the Y term of an $X \rightarrow Y$ associative pair in the Ryan and Mink studies suggested that associative relationships might account for the variation in G word response in the present study. If G word response varied with the strength of associative relationships between G words and L list words, connotative similarity and association might be assumed to have functioned as dual determinants of the observed generalization. If substantial associative relationships were found between L list words and the "functional" group of G words but none were found between L list words and the "nonfunctional" group of G words, the associative factor might plausibly be considered the major determinant of the observed generalization.

Since the Minnesota norms contain associative data for only a small part of the words used in the present experiments, free association data for these words were obtained from a sample of 128 college students. The format and instructions of the association study were identical to those used in the Minnesota study. The order of presentation of stimuli was rotated with respect to classification in the generalization studies: an A word, a B word, an Experiment 1 F word, an Experiment 1 C word, an M word, an N word, an Experiment 2 F word, and an Experiment 2 C word. No G word appeared in the list until all L list words had appeared. This was done to control for the possibility that G words would occur as associative responses to L list words with a spuriously high frequency because of recency of prior occurrence as stimuli. (Storms—1956—found that recent occurrence of Y as a stimulus results in higher frequencies of Y as a free association response.) By necessity, this arrangement of the list is biased toward spuriously high frequencies of L list words as associative responses to G words.

Results

L word → G word associations (Mink paradigm). Mink found associative habits between words were relevant to generalization only when the learning list words were the stimulus terms (X) and the test list words were the response terms (Y) of associative pairs ($X \rightarrow Y$). Table 5 shows the associative data relevant to the possibility of Mink-paradigm associative generalization in Experiments 1 and 2. The left-hand columns show all L list words which associatively elicited the G word shown in the right-hand column, and the percentage of the 128 subjects who gave the G word as an associative response to each L list word. The percentages are not independent. G word GIRL, for instance, was the associative response to L list word PRETTY for 23% of the subjects, the response to WOMAN, for 12% of the subjects, etc., with no constraint against the same subject responding GIRL to more than one of these L list words.

The data indicate that Mink-paradigm generalization was a negligible factor in the outcomes of Experiments 1 and 2. Even if the nonindependent response frequencies are cumulated, only one G word of the 16 (GIRL) occurs frequently enough as an associative response of L list words to determine generalization of the kind Mink observed. Most of the G words with high response frequencies in the generalization experiments have zero or very small frequencies as associative responses to L list words.

G word → L list word associations (Shipley-Lumsdaine paradigm). Application of the Shipley-Lumsdaine paradigm of

TABLE 5

PERCENT INCIDENCE OF G WORDS AS ASSOCIATIVE RESPONSES TO L LIST WORDS

(Associations Relevant to Mink Paradigm)

Experiment 1			Experiment 2		
L list word(s) which elicit G word and percentage elicitation of G word by L list word		G word as Response	L list word(s) which elicit G word and percentage elicitation of G word by L list word		G word as Response
		A" words			*M"* words
PRETTY	23.4	GIRL	STUDY	.8	EFFORT
WOMAN	11.7		STEAL	.8	LIFT
LOVELY	10.9				
BODICE	4.7		(No elicitor)		INCOME
(No elicitor)		CHARMING	(No elicitor)		UP
(No elicitor		FRAGRANT			
(No elicitor)		SONG			
		B" words			*N"* words
TROUBLE	2.3	DANGER	STEAL	11.7	THIEF
AFRAID	1.8		CROOKED	3.9	
(No elicitor)		FRIGHTFUL	BAD	1.6	CRIMINAL
			CROOKED	.8	
(No elicitor)		PAIN	STEAL	.8	
			MONEY	.8	
(No elicitor)		HURT	FEAR	3.1	HATE
			(No elicitor)		FRAUD

association-determined generalization to the context of the current experiments requires associative habits of form (G word → L list word). The application of the paradigm can be clarified by substituting *Y*, the notation adopted for the response term of an associative pair for "L list word," and *X*, the notation for the stimulus term of an associative pair for "G word." If associative habits determine generalization in the Shipley-Lumsdaine paradigm (Figure 4), a lever response learned to an L list word (*Y*) should generalize to a G word (*X*) when (G word → L list word) associations are of sufficient strength.

Table 6 shows the percentage of the 128 subjects who responded associatively with L list words when G words were used as stimuli. No control word elicited an L list word, and no F words were elicited. The percentages for different L list words elicited by the same G word are independent, since each G word appeared only once and elicited a single response from each subject.

Associative relationships of some G and L words in the (G word → L list word) direction are considerably greater than those in the reverse direction. This is particularly true if response percentages for the several L list words elicited by a single G word are

combined.[6] Four of the 16 G words (CHARMING, GIRL, INCOME, THIEF) elicit one or another L list word in more than 20% of the subjects. Two more G words (CRIMINAL, FRAUD) elicit some L list word in more than 10% of the subjects. All 6 of these G words fall in the "functional" category in regard to generalization. If the 6 "nonfunctional" G words are compared with the 10 "functional" G words, frequency of (G word → L list word) associations is significantly greater for the latter at the .05 confidence level.[7]

[6] Such cumulation is *statistically* legitimate in view of the independence of the percentages. Whether a set of relatively low-frequency associations can combine to yield sufficient associative "strength" to determine generalization is not known. Such combination seems logical, however, since it is a matter of indifference in the present context *which* L list word is an associate of a G word, and since the "strength" of a given associative pair is by definition maximal in any individual subject for whom X associatively elicits Y, irrespective of the infrequency of $(X \rightarrow Y)$ in group data.

[7] Mann-Whitney U test, two-tailed.

TABLE 6

PERCENT INCIDENCE OF L LIST WORDS AS ASSOCIATIVE RESPONSES TO G WORDS

(Associations Relevant to Shipley-Lumsdaine Paradigm)

Experiment 1		Experiment 2	
G word as Stimulus	L list word as response to G word and percentage elicitation of L list word	G word as Stimulus	L list word as response to G word and percentage elicitation of L list word
A″ words		*M″* words	
CHARMING	WOMAN 12.5 PRETTY 7.0 LOVELY 5.4	INCOME	MONEY 55.4
		EFFORT	PROGRESS .8
GIRL	WOMAN 12.5 PRETTY 10.2	UP	(None)
		LIFT	(None)
SONG	PRETTY 1.6		
FRAGRANT	(None)		
B″ words		*N″* words	
FRIGHTFUL	AFRAID 5.4 WOMAN .8 LOVELY .8	THIEF	STEAL 33.6 BAD 4.7 MONEY 1.6 AFRAID .8 CROOKED .8
DANGER	AFRAID 3.9 TROUBLE 3.1	CRIMINAL	BAD 10.9 STEAL 1.6
HURT	(None)		
PAIN	(None)	FRAUD	STEAL 9.4 BAD 3.9 MONEY 1.6 STUDY .8
		HATE	(None)

Three factors must be considered in evaluating the possibility that (G word → L word) associations (Shipley-Lumsdaine paradigm) have a relationship to the generalization frequencies of the G words.

1. The fact that the observed associative frequencies may be spuriously elevated by the recency factor (Storms, 1956) may account for the higher association values in Table 6 compared with those in Table 5. This could vitiate an interpretation of G word response in terms of association-determined generalization. The *differences* in observed association values of the functional and nonfunctional G words in Table 6 cannot, however, be accounted for by the recency factor.

2. The possibility of spurious associative frequencies due to recency, does, however, contribute to the cogency of a second consideration. The association frequencies of (G word → L word) relationships are with two exceptions below the level Mink (1957) found necessary for generalization. This is true even when response percentages of different L list words are cumulated. If the observed (G word → L word) associative frequencies are spuriously high because of the recency factor, the possibility that any of the observed generalization can be attributed to associative factors is, of course, further reduced. In any instance, it is clear that the associative frequencies are not sufficient to explain all of the generalization, since several of the functional G words do not associatively elicit L list words.

3. The most important consideration is Mink's carefully replicated finding that generalization does not occur on the basis of associative relationships employed in the Shipley-Lumsdaine paradigm. Unless later evidence reverses Mink's findings, it does not appear plausible to attribute the differences in G word generalization in the present experiments to associative habits in the (G word → L word) direction.

The absence of associative relationships appropriate to the Mink paradigm and the above considerations concerning associative relationships relevant to the Shipley-Lumsdaine paradigm suggest that the association data do not account for the differences in G word response in Experiments 1 and 2. The observed relationship between (G word → L word) associations and generalization do, however, indicate the value of further investigation of the Shipley-Lumsdaine associative mediation paradigm and the need for control of the recency factor in gathering associative data for clarification of the generalization problem.

The Effect of Change of Verbal Context

The outcome of Experiments 1 and 2 suggested two possible relationships between the cluster predominance phenomenon and the wide variability in response to individual G words.

1. Predominance of one cluster over the other in the subject's attention during learning and/or generalization might be intrinsic to the two-cluster design. If this were so, variation in response to individual G words might be in part attributable to "suppression" of response to the G words in the subordinated cluster. Five of the six nonfunctional G words in the first two experiments were from the subordinated clusters. Variability in response to individual G words within the same cluster is not, of course, explained by this hypothesis.

2. The cluster predominance phenomenon may be an artifact of differences in response potentialities of individual G words. G words which are nonfunctional due to unknown factors not controlled in the determination of connotative similarity could have accumulated by chance in disproportionate numbers in the two "weak" subclusters. The between-cluster differences in generalization might reflect only differences in individual word response potentiality.

If Hypothesis 1 were correct, the predominance effect should occur in any two-cluster experiment, regardless of the response potentialities of individual G words. If two clusters containing apparently potent G words (for example, the two dominant

clusters A and N) were paired in the same experiment, response to one subcluster of G words should be substantially less than response to the other. Individual stimuli in the "suppressed" cluster would elicit less response than before, and some G words might be expected to shift from functional to nonfunctional status. If two previously subordinate clusters were paired in an experiment, one of the clusters would be expected to become dominant, with an increase in individual G word response and some shift from nonfunctional to functional status of G words.

If Hypothesis 2 were correct, no cluster difference effect would be expected when two subclusters of potent G words appear in the same experiment. Pairing of two previously weak clusters would similarly yield little or no cluster difference in generalization. Response frequencies of individual G words would be expected to remain at approximately the same levels observed in the original experiments.

Experiments 3 and 4 were designed to evaluate these alternative explanations of the cluster effect. The two previously dominant clusters (A and N) were paired in Experiment 3, and the two previously subordinate clusters (B and M) were used together in Experiment 4. The Experiment 1 control words were used in Experiment 3, since all D values for control word-experimental word connotative comparisons exceeded the previously adopted standard of 6.00. The Experiment 2 control words met the same standard with respect to all Experiment 4 experimental words, and were used as control words in Experiment 4. Four filler words were selected for each experiment so that neither experimental cluster was favored in connotative similarity. Some F words were the same used in the earlier experiments, and one was used for both Experiments 3 and 4. The verbal items used in the two new experiments and the altered context of the experimental clusters can be seen in Tables 7 and 8. Procedures for conducting the experiments were identical to those of the first two experiments.

Results

Experiment 3. Table 7 shows the outcome of Experiment 3. Response to A'' words was 14.2% of the total possible response, in contrast to the 29.6% response to the same words in Experiment 1. Response to N'' words was 15.4%, in contrast to 35.4% response to these words in Experiment 2. Control word response was 4.2% in Experiment 3, compared to 7.5% response to the same words in Experiment 1. $\bar{G}_T$ was 5.10 ($p < .001$); $\bar{G}_A$ was 4.80 ($p < .001$); $\bar{G}_N$ was 5.40 ($p < .001$); $\bar{D}_G$ was .30 ($p > .10$).

Experiment 4. Table 8 shows the outcome of Experiment 4. Response to B'' words was 20.0% of the total possible response, in contrast to 14.0% response to these words in Experiment 1. Response to M'' words was 8.1%, in contrast to 19.2%

TABLE 7

PERCENTAGE RESPONSE TO INDIVIDUAL WORDS: TEST TRIAL, EXPERIMENT 3

L list words	% response	Words unique to T list	% response
A'		*A''*	
WOMAN	97	GIRL	25
BODICE	93	CHARMING	19
PRETTY	77	FRAGRANT	11
LOVELY	67	SONG	2
		All *A''* words	14.2
N'		*N''*	
BAD	72	HATE	27
NASTY	61	THIEF	21
STEAL	50	CRIMINAL	9
CROOKED	42	FRAUD	5
		All *N''* words	15.4
F		Control	
CITY	82	FAT	14
ART	75	PLAIN	5
RAPID	70	SLOW	4
AFRAID	46	SNAIL	3
		DIM	3
		TRUNK	2
		SQUARE	1
		LONG	0
		All C words	4.2

TABLE 8

PERCENTAGE RESPONSE TO INDIVIDUAL WORDS: TEST TRIAL, EXPERIMENT 4

L list words	% response	Words unique to T list	% response
B'		*B''*	
SCORCHING	98	DANGER	27
SCALDING	96	FRIGHTFUL	22
MAD	79	PAIN	20
TROUBLE	65	HURT	11
		All *B''* words	20.0
M'		*M''*	
PATRIOT	86	EFFORT	12
MONEY	78	INCOME	11
PROGRESS	75	LIFT	5
STUDY	41	UP	4
		All *M''* words	8.1
F		Control	
BOX	72	MOLD	14
BLOCK	66	STOUT	7
AFRAID	44	SLEEP	5
LIGHT	35	CALM	2
		PIG	2
		SOFT	2
		ROUND	1
		DUSKY	1
		All C words	4.1

response to the same words in Experiment 2. Control word response was 4.1% in Experiment 4, compared to 9.4% response for the same words in Experiment 2. $\bar{G}_T$ for Experiment 4 was 4.80 ($p < .001$); $\bar{G}_B$ was 7.75 ($p < .001$); G_M was 2.10 ($p < .05$); $\bar{D}_G$ was 2.85 ($p < .01$).

The data again confirm the main generalization effect. All four clusters of G words demonstrated significant generalization. Response for many individual G words is substantially greater than response to control words. However, the most striking result of the context-change experiments was the marked shift in response levels of individual G words, G word clusters, and control words, relative to the levels elicited by the same stimuli in the initial experiments. Three of the four G word clusters elicited less than half the response frequencies observed earlier, while response to the remaining cluster increased. One set of control words elicited less than half the previous response, and the other set evoked only slightly more than half. Pronounced shifts in response to some individual G words can be seen by comparing Tables 3 and 4 with Tables 7 and 8.

The decrease in response frequencies for all classes of test list words was without apparent explanation. Change of context had been presumed possibly relevant to change in relative amounts of response for clusters of G words, but had not been expected to reduce the response to *both* clusters, as occurred in Experiment 3, or to materially affect response to the control words.

The possibility of some kind of reduction of the general activity level of the subjects was considered: The subjects for Experiments 3 and 4 were drawn from the same psychology classes as those used in the initial experiments. They were motivated to participate by the same incentive of two points final examination credit. Since the summer weather was somewhat uncomfortable, temperature and humidity data for the times of the two sets of experiments were examined, but no differences were found. An hypothesis of lowered activity level is in addition inconsistent with two findings: response to one group of G words (*B''*) *increased* in the second set of experiments, and response to L list words on the T list remained relatively constant in the two sets of experiments.[8]

Response to Individual G Words

A test of the functional or nonfunctional status of the G words as elicitors of generalized response was conducted for the data of Experiments 3 and 4 in the same manner as before. The distributions of response frequencies for the control words in the two context-change experiments were compared and pooled when no significant differences in means or variances were found.

[8] One group of L list words (*N'*) decreased in response frequencies, but the other three groups (*A'*, *B'*, and *M'*) evoked comparable response levels in both the initial and context-change experiments.

G words with response frequencies greater than that of the most popular control word were tentatively considered "functional," the remaining G words "nonfunctional." The least popular functional G word exceeds the control word mean by 3.47 sigma, a deviation expected 1 time in 1,000 in a distribution with the parameters of the control word response distribution. The most popular nonfunctional G word exceeds the control word mean by only 1.93 sigma, a deviation which is not statistically significant at the .05 level. Classification of G words as functional or nonfunctional generalization stimuli in the context-change experiments thus appears satisfactory, although it is not quite as distinct as the classification on the basis of the data from the original experiments. Table 9 shows the classifications of the 16 G words in the two contexts. Nine words retain their previous status in the altered context, but seven change classification.

Stability of the functional status of more than half of the G words in the new context appears to confirm the hypothesis of stable differences in the response potentialities of at least some of the individual generalization stimuli. This tends to substantiate Hypothesis 2 above. The continued wide variability of response to G words within subclusters also tends to confirm Hypothesis 2. Although one cluster of G words yielded significantly more generalization than the other in Experiment 4, the cluster effect fails in Experiment 3, a finding more consistent with Hypothesis 2 than with Hypothesis 1. However, in spite of these trends in the data, the unstable response levels for clusters as well as for individual words indicated there was no unequivocal choice between the alternatives which Experiments 3 and 4 had been designed to resolve. Context change appeared to be a factor which introduced a new component of variability in the response frequencies for some of the G words, but the unexplained and large variation in over-all response levels suggested that unreliability of the experimental procedures might account for the changes. Confirmation of the role of context change in the variability of individual word response frequencies and resolution of the interpretation of the cluster phenomenon appeared to require further verification of the reliability of the relative differences in response frequencies of individual generalization stimuli.

TABLE 9

COMPARISON OF FUNCTIONAL STATUS OF G WORDS: EXPERIMENTS 1 AND 2 VS. EXPERIMENTS 3 AND 4

Experiment 3 or 4	Experiment 1 or 2	
	Functional	Nonfunctional
Functional	(*A″*) GIRL (*A″*) CHARMING (*B″*) FRIGHTFUL (*N″*) THIEF (*N″*) HATE	(*B″*) DANGER (*B″*) PAIN
Nonfunctional	(*A″*) FRAGRANT (*M″*) EFFORT (*M″*) INCOME (*N″*) CRIMINAL (*N″*) FRAUD	(*A″*) SONG (*B″*) HURT (*M″*) LIFT (*M″*) UP

REPLICATION OF THE INITIAL EXPERIMENTS

Exact replications of the initial generalization experiments were conducted to determine the reliability of the absolute and relative response frequencies for clusters of generalization stimuli and for individual G words. Experiment 5 duplicated the materials, list sequences, and procedures of Experiment 1; Experiment 6 repeated Experiment 2.

Results

Tables 10 and 11 show the outcomes of the replication experiments.

Experiment 5. Response to *A″* words was 18.3% of the total possible response, in contrast to 29.6% in the first experiment and 14.2% in the context-change experiment. Response to *B″* words was 12.1% in replication, in contrast to 14.0% in the first experiment and 20.0% in the changed

TABLE 10

PERCENTAGE RESPONSE TO INDIVIDUAL WORDS: TEST TRIAL, EXPERIMENT 5

L list words	% response	Words unique to T list	% response
A'		*A''*	
BODICE	100	CHARMING	35
LOVELY	76	GIRL	18
WOMAN	76	FRAGRANT	16
PRETTY	62	SONG	04
		All *A''* words	18.3
B'		*B''*	
SCORCHING	87	FRIGHTFUL	19
SCALDING	82	DANGER	14
MAD	77	PAIN	11
TROUBLE	56	HURT	04
		All *B''* words	12.1
F		Control	
SUCCESS	98	SQUARE	09
FOOD	72	TRUNK	05
CAR	69	PLAIN	04
AFRAID	51	DIM	03
		LONG	02
		FAT	02
		SNAIL	02
		SLOW	01
		All C words	3.6

context. The control words yielded 3.6% response, compared to the 7.5% response observed in the first experiment and the 4.2% in the third experiment. $\bar{G}_T$ for Experiment 5 is 5.55 ($p < .001$). $\bar{G}_A$ is 7.05 ($p < .001$) and $\bar{G}_B$ is 4.05 ($p < .02$). The mean cluster difference score $\bar{D}_G$ is 1.50 ($p > .05$).

Experiment 6. Response to *M''* words was 17.1% of the total possible response, in contrast to 19.2% response to these words in the second experiment and 8.1% response in the context-change experiment. Response to *N''* words was 22.9% in replication, compared to 35.4% in the second experiment and 15.4% in the changed context. Control words received 5.2% response in Experiment 6, compared to 9.4% in Experiment 2 and 4.1% in Experiment 4. $\bar{G}_T$ for Experiment 6 is 7.10 ($p < .001$). $\bar{G}_M$ is 5.70 ($p < .001$), and $\bar{G}_N$ is 8.50 ($p < .001$). $\bar{D}_G$ is 1.40 ($p > .05$).

Response to Individual G Words

The distributions of response frequencies of the two sets of control words in the replication experiments were compared, found not significantly different in means or variances, and pooled. Response frequencies of 4 of the 16 G words (SONG, HURT, UP, LIFT) fall within the range of the pooled control word distribution and deviate less than one sigma from the control word mean. These four words are consistently nonfunctional in all three sets of experiments. G word PAIN, classified nonfunctional in the initial experiment but functional in the changed context falls outside the pooled control word distribution in the replication experiment and deviates from the control word mean by 2.22 sigma.

TABLE 11

PERCENTAGE RESPONSE TO INDIVIDUAL WORDS: TEST TRIAL, EXPERIMENT 6

L list words	% response	Words unique to T list	% response
M'		*M''*	
PATRIOT	95	EFFORT	30
PROGRESS	72	INCOME	29
STUDY	67	UP	07
MONEY	67	LIFT	02
		All *M''* words	17.1
N'		*N''*	
BAD	74	HATE	27
STEAL	61	THIEF	24
CROOKED	57	FRAUD	21
NASTY	54	CRIMINAL	19
		All *N''* words	22.9
F		Control	
ABRUPT	88	STOUT	09
CITY	72	SOFT	08
FEAR	71	ROUND	07
BLOCK	61	CALM	07
		MOLD	04
		SLEEP	03
		PIG	02
		DUSKY	0
		All C words	5.2

This word thus appears of borderline functional status in the replication experiment, and may be classified nonfunctional if the .01 level of significance is adopted. The remaining 11 G words deviate from the control word mean by more than three sigma units, and are clearly classifiable as functional in the replication experiments.

Comparison of Initial, Replication, and Context-Change Data

Table 12 summarizes the G word data for the three sets of experiments. Response percentages for each G word in initial (I), context-change (X), and replication (R) experiments, and percentage response in all experiments (T) are shown in the first column. Column 2 shows the functional status of each G word in each experiment. Data in the remaining columns are discussed in the following section.

Several conclusions can be drawn from comparison of the response data from the three sets of experiments:

1. The experimental method reliably demonstrates the hypothesized connotatively-determined generalization for clusters of generalization stimuli. Total response to G words and response to subclusters of G words significantly exceeds control word response in every instance in the six experiments.

2. Differences in relative levels of response for individual G words in a constant verbal context are reliably demonstrated. Depending on the choice of significance level for classification of the "borderline" word PAIN in the replication experiment, either 14 or 15 of the 16 G words are consistently classified in the initial and replication studies as to whether they function or fail to function as generalization stimuli. The product-moment correlation between frequency of response to the 16 G words in the initial and replication experiments is .89 ($p < .01$).

3. The importance of the role of verbal context in determining response frequencies and functional status of some individual G words appears reliably established. Only 9 of 16 G words are consistent in functional status in an altered verbal context, in contrast to 14 or 15 in the exact replication study. The correlations of response levels of G words in the initial experiments and the context-change experiments is .41, and the value for the context-change data vs. the replication data is .43. Neither value is large enough for statistical significance, although the level of correlation suggests some stability of response level irrespective of context.

4. Connotative meaning similarity to learning list words is always a determinant of generalization in the case of some stimuli, and never a determinant in the case of some other stimuli. Five G words are consistently functional (marked CF) in Table 12, Column 2, and four G words are consistently nonfunctional (marked CNF).

5. The experimental method is considerably less reliable for the measurement of absolute levels of generalized response than for relative levels. The distinct drop in frequency of response to Subclusters *A''* and *N''* which occurred in the context-change experiments also occurs in the replication experiments. Response to control words was similarly much lower in both context-change and replication experiments than in Experiments 1 and 2. The significant cluster predominance effect which appears in Experiments 1, 2, and 4 does not appear in the other three experiments, and appears interpretable as an artifact of fluctuation of response levels for subclusters due to unknown factors. The distinctive shifts in absolute response levels to subclusters of G words and to individual G words indicate the necessity of control words as baselines for response frequency in verbal experimentation.

Usage Frequency, Meaningfulness, and Denotation in Relation to Level of Generalized Response

The association study indicated that associative habits could not account for the differences in G word response in the initial experiments. Inspection of the association

TABLE 12

SUMMARY OF G WORD DATA

	G Word	1 Percentage Response I	X	R	T	2 Functional Status* I	X	R	T	3 Thorndike-Lorge Frequency	4 Meaningfulness	5 Proportion of "Similar" Judgments Part I Data	6 Proportion of "Similar" Judgments Part II Data	7 Primary and Secondary Associates
A''	CHARMING	57	19	35	37	+	+	+	CF	31	6.089	.80	.92	(NICE) (WOMAN)
	GIRL	32	25	18	25	+	+	+	CF	*AA*	6.498	.88	.90	(BOY) (WOMAN)
	FRAGRANT	25	11	16	17	+	0	+		14	5.713	.13	.13	(SMELL) (FLOWER)
	SONG	4	2	4	3	0	0	0	CNF	*AA*	5.543	.17	.05	(SING) (MUSIC)
B''	FRIGHTFUL	28	22	19	23	+	+	+	CF	10	5.659	.44	.62	(AWFUL) (AFRAID)
	DANGER	9	27	14	17	0	+	+		*AA*	5.492	.48	.62	(FEAR) (STOP)
	PAIN	12	20	11	14	0	+	0		*AA*	4.979	.50	.38	(HURT)
	HURT	7	11	4	7	0	0	0	CNF	*AA*	5.437	.57	.38	(PAIN)
M''	EFFORT	38	12	30	27	+	0	+		*AA*	5.558	.87	.90	(WORK) (TRY)
	INCOME	21	11	29	20	+	0	+		46	6.213	.45	.56	(MONEY) (TAX)
	UP	10	4	7	7	0	0	0	CNF	*AA*	4.604	.32	.36	(DOWN) (SKY)
	LIFT	7	5	2	5	0	0	0	CNF	*AA*	4.956	.32	.18	(UP) (HEAVY)
N''	HATE	41	27	27	32	+	+	+	CF	*A*	6.943	.25	.17	(LOVE) (DISLIKE)
	THIEF	38	21	24	28	+	+	+	CF	28	6.436	.65	.66	(STEAL) (CROOK)
	CRIMINAL	32	9	19	20	+	0	+		31	6.695	.63	.77	(THIEF) (BAD)
	FRAUD	30	5	21	19	+	0	+		10	5.486	.47	.40	(CHEAT) (STEAL)

Note.— I = Initial Experiments (1 or 2).
X = Context-Change Experiments (3 or 4).
R = Replication Experiments (5 or 6).
T = Total, all experiments.
* At .01 significance level for deviation from control word mean.

data for the additional (L word → G word) and (G word → L word) associations of possible relevance in the context-change experiments[9] indicated no frequencies as great as 5%. The original association data and this additional data appear to quite definitely rule out Mink paradigm or Shipley-Lumsdaine paradigm associative generalization in the context-change experiments. Three additional variables which might account for differences in G word response levels are considered in this section.

Usage Frequency

Column 3 of Table 12 shows the Thorndike-Lorge (1944) usage frequency values of the G words. Numbers refer to usage frequency per 1,000,000 words. *A* designates words with frequencies between 50 and 99 per 1,000,000, *AA* words with frequencies of 100 or more per 1,000,000. The frequency values do not appear to systematically relate to amount of generalized response or to the functional-nonfunctional classification. It may be noted that no really infrequent words were used.

Meaningfulness

Column 4, Table 12 shows *connotative meaningfulness* values of the G words, defined by

$$M = \sqrt{\frac{\Sigma}{i} d_i^2}$$

where i is the number of Semantic Differential scales used in measuring the connotative meaning of a word and d_i is the deviation of the scale value of a word from the midpoint of the scale. This index measures the deviation of the Semantic Differential profile of a word from that of a completely neutral or "connotatively meaningless" stimulus. When M is large, scale values are extreme, and the stimulus may be considered vivid or powerful in connotation. Construct validity for M is suggested by its high correlation (.71) with Noble's association-based meaningfulness index (Jenkins & Russell, 1956). The functional potency of the mediation process $r_m s_m$ associated with a word might plausibly be related to the value of M. If this were so, words with large M values should have a relatively great potentiality for eliciting an instrumental response (cf. Figure 1).

The relationship between the M values of the G words and level of generalized response in the present experiments substantiates these assumptions. The Pearsonian correlation of M with total generalized response for the 16 G words is .66 ($p < .01$). There is no overlap of the M values of the five consistently functional and the four consistently nonfunctional G words. The mean M value for the CF words is 6.32, for the CNF words 5.13 ($p < .01$, t test). The data indicate that the meaningfulness factor is important in the observed differences in generalization to individual stimuli.

Judged Denotative Similarity

Inspection of the G words with consistently low response frequencies suggested that some of these words are less similar denotatively to learning list words of the same connotative cluster than are the G words which yielded high generalization frequencies. *A''* word SONG, for instance, which consistently failed to elicit generalization, does not seem to partake as much of a denotation which might be termed "femininity," as do all of the more potent *A''* words (CHARMING, GIRL, FRAGRANT) and the learning list words of the *A* cluster (WOMAN, BODICE, PRETTY, CHARMING). Similarly, *N''* words UP and LIFT, consistently nonfunctional, do not appear as denotatively similar to their learning list connotative counterparts (PROGRESS, PATRIOT, STUDY, MONEY) in terms of the denotation "social-value-related concept" as do the more successful *N''* generalization stimuli (EFFORT and INCOME).

[9] These are "Mink paradigm" associations of the form *A'* word → *N''* word, *N'* word → *A''* word, *B'* word → *M''* word, and *M'* word → *B''* word, and "Shipley-Lumsdaine paradigm" associations of the form *N''* word → *A'* word, *A''* word → *N'* word, *B''* word → *M'* word, and *M''* word → *B'* word.

To test the possibility that differences in denotative relationships relate to the observed differences in generalization, 24 male and 15 female college students were asked to judge the denotative similarity of each G word in relation to the four relevant, connotatively similar L list words. Appendix C illustrates the two kinds of forms on which the judgments were recorded. In Part I of the study, 16 pages like that shown in Appendix C-I were used. Subjects were asked to rank the four G words of a subcluster according to amount of denotative similarity to one of the revelant L list words. Denotation was defined at several points in the instructions as the "thing to which the word refers." Each set of G words was ranked against each of the four relevant L list words, although pages pertaining to Clusters *A*, *B*, *M*, and *N* were rotated so that words from each cluster appeared only once in every four pages.

Part II of the study used four pages like that shown in Appendix C-II. Four G words from a given cluster appeared below as before, and the entire learning list subcluster relevant to the G words appeared above. Subjects were asked to consider the L list subcluster at the top of the page until they had an idea of the "thing" these words referred to *in common*, and then to rank the G words as to similarity in denotation to this common meaning. All 39 subjects made all 16 sets of rankings in Part I and all 4 sets in Part II, with the exception of a few omitted judgments.

Although the subjects when questioned after the judgments indicated they had no difficulty in understanding the judgment task or the definitions of "denotation" and "denotative similarity," there is of course no guarantee that denotative similarity was in fact the quality being judged. Connotative similarity, word association, or other factors may have influenced the judgments. Accordingly, the data are termed "judged denotative similarity," and are limited by these considerations.

Column 5 of Table 12 shows the proportion of Ranks 1 and 2 earned by each G word in all judgments made in Part I. Each proportion is determined by approximately 156 judgments (39 subjects × 4 judgments for each G word). The proportion is an index of the relative number of "similar" rankings of each G word in relation to the relevant L list words. The rank order correlation of the proportion of "similar" denotative judgments and the total amount of generalized response across all 16 G words is .51 ($p < .05$). If denotation is an important determinant of generalized response, as the correlation seems to indicate, the five CF words might be expected to receive predominantly (more than half) "similar" judgments, while the four CNF words should receive predominantly dissimilar judgments. The proportions are in the expected direction in six of nine cases, a trend consistent with the hypothesis but statistically insignificant.

Column 6 of Table 12 shows the proportion of 1 and 2 ranks ("similar") earned by each G word in Part II of the judgment study. Each proportion is determined by 39 judgments, since each G word was ranked against the relevant L list subcluster only once by each subject. These denotative judgments are more closely related to the amount of generalized response than the Part I judgments. Clusters *A''* and *M''* show perfect correspondence in intracluster rank orders of total generalized response frequency and proportion of similar judgments. The correspondence of these rankings in Cluster *B''* is as close as tied ranks on the denotative variable permit. No intracluster correspondence is observed, however, for Cluster *N''*. The rank order correlation of amount of generalized response and proportion of similar judgments across all 16 G words is .66 ($p < .01$). The tendency for more than half "similar" judgments for CF words and less than half "similar" judgments for CNF words is in the expected direction in eight of nine cases, a trend significant at the .05 level (sign test).

Both kinds of judgment data appear to confirm the hypothesis that denotative meaning factors not measured by the Semantic Differential play an important role in the obtained generalization. One hypothesis suggested by the data is that some degree

of denotative similarity is a necessary condition for generalization. If it can be assumed that connotation and denotation are related, it may be that the Semantic Differential measurements select words with the necessary denotative similarity in a majority of instances, but fail to do so in some cases.

The Part II judgment data were obtained in a context most similar to that of the experiments, in that a *cluster* of L words was judged in relation to a cluster of G words. The context factor appears to be one which requires control in judgments of this kind as well as in generalization proper, since the Part I judgment data, obtained in a somewhat different context, do not predict generalization frequencies as well as the Part II data.

Another aspect of the hypothesis concerning the role of denotative similarity in generalization concerns the primary and secondary associates of the G words. These were determined by the most frequent and next most frequent response in the free association data, and are shown in Column 7 of Table 12. The primary associate is first in each case. When there was no secondary associate with a response frequency of 5% or greater, none is shown. The data allow examination of the possibility that "functional" G words elicit frequent associates which are similar denotatively to relevant L list words, while "nonfunctional" G words elicit associates more remote in denotation in relation to the L list words. Three of the four CNF words (SONG, LIFT, UP) have both primary and secondary associates whose denotation appears remote from the denotation of the relevant L list words. In contrast, none of the five CF words have both primary and secondary associates which appear foreign to the pertinent L list words denotatively, although in one case the primary associate is a word opposite in meaning to an L list word. Judgment of similarity of denotation is here a priori, so the comparison is suggestive only. It appears, however, that this associational index of similarity, which assumes that the associates of a word are indices of its "meaning" (Noble, 1952) lends some additional confirmation to the hypothesis that a meaning similarity factor other than scaled connotative meaning is a determinant of the differences in generalized response.

Discussion

The results of the series of experiments and the supplementary studies may be examined with reference to the main hypotheses of the research, which were specified in the initial section. The consistent elicitation of generalized response by groups of word stimuli selected for measured connotative similarity to learning list words appears to substantiate the hypotheses. The fact that generalization occurs between stimuli selected solely on the basis of quantitative comparisons of independently-obtained meaning profiles for words suggests the usefulness of the conceptualization of meaning as a mediation process (Hypothesis *a*) and of Osgood's operation formulation of meaning and meaning similarity (Hypotheses *b* and *c*).

The hypotheses of the study did not, however, predict the variability in generalized response which was observed. Aside from the problem of instability in response levels of groups of stimuli and individual words (which cannot be accounted for by any variables examined here), several considerations emerge from the data which require modification of the conclusion that the major hypotheses are confirmed.

One kind of consideration has to do with factors other than measured connotative meaning which are related to the observed generalization. The verbal context of those stimuli which are crucial to the generalization paradigm (Figure 2) appears without question to be one of these factors. The pronounced shifts in generalized response in the context-change experiment, which exceed the shifts in the replication studies, indicate the context factor is powerful enough to substantially alter the effect of the connotative meaning variable. The data indicate that context is likely to be a complicating factor in any verbal research of this kind in which more than a very small number of stimulus elements are

subjected to experimentation with the same subjects at the same time.

The vividness of connotative meaning as measured by M is a factor determinable by the Osgood technique but which was not assumed relevant to generalization until after the data had been gathered. This factor also appears to be a major one in determining level of generalization, but need not be a serious problem in further research because of the relative ease with which it could be controlled.

Because of difficulty in obtaining "pure" judgments of denotative relationships, the data pertaining to the role of denotative similarity in the present findings are not entirely satisfactory. Nonetheless the denotative factor appears, not surprisingly, to be one which affects mediated generalization. The Osgood meaning measurement does not purport to measure denotation, and it is difficult to imagine any practicable extension of the polar scaling technique to cover the almost infinite scope of differences in denotative meaning. Some control of denotation in further work on measured connotative meaning may, however, be necessary.

A second kind of consideration, more serious in its implications for the interpretation of the data in relation to the major hypotheses, is the fact that some stimuli selected for connotative similarity to learning list words unvaryingly failed to elicit any generalized response. For at least 4 of the 16 G words, high similarity of connotative meaning in relation to learning list words did not result in generalization. If a stringent scientific logic is adopted, even a single negative instance of an affirmative generalization (or set of generalizations) demands refutation of the generalization. If the individual G word is taken as a unit of evidence, 25% of the hypothesized instances of generalization may be said to have failed to materialize in any of three independently conducted sets of experiments. The question may be raised whether any of the hypotheses may be considered confirmed in view of this amount of negative evidence.

However, the contemporary viewpoint on the use of networks of hypotheses and "open" definition of theoretical constructs (Braithwaite, 1953; Pap, 1953), urges flexibility in modification of a set of hypotheses which appear in the main confirmed by evidence, but which are also partially at variance with the evidence. An entire set of hypotheses need not be considered refuted by instances of negative evidence, but instead a decision can be made as to the point in the theoretical framework which may be most strategically modified. In this view, any of the Hypotheses a, b, or c could be considered inadequate on the basis of the negative evidence, and correspondingly modified. The remaining hypotheses would be retained on the basis of substantiation by the positive findings. The mediation hypothesis (Hypothesis a) might, for instance, be treated as disconfirmed, at least for some of the generalization stimuli. Alternatively, R_{SD} might be considered to have been shown inadequate as an index of the mediator (Hypothesis b). The Semantic Differential might be regarded as a measurement which overlaps partially with the elements of the mediator $r_m s_m$, but which is not sufficiently similar for precise measurement. Or it might be hypothesized that the mediation hypothesis and the meaning measurement technique are valid, but that D adequately measures connotative similarity for only part of the interword comparisons (Hypothesis c).

Modification of Hypothesis b seems the most plausible way to interpret the negative evidence from the present study. The mediation hypothesis seems the best confirmed of the three on the basis of previous research on secondary generalization, and it seems reasonable to interpret it as additionally substantiated by the main body of the present evidence. The hypothesis that R_{SD} is a suitable operational reduction of the mediation construct seems, in contrast, only partly confirmed by the present evidence. The positive evidence is interpretable as lending some "construct validity" (Cronbach & Meehl, 1955) to the Semantic Differential as a measure of the representational mediation process, but the nega-

tive evidence suggests that R_{SD} cannot be considered isomorphic with this process. The role of denotative meaning and of meaningfulness in determining generalization strengthen this interpretation. It is possible that better control of extraneous factors would function to increase the validity of R_{SD}, although consistent absence of generalization to the four CNF words suggests an upper limit to the increase.

Explanation of negative evidence by modification of Hypothesis *c* is also a possibility. Alternative methods of measuring similarity (for instance, focus on factorial dimensions rather than profile differences with all scale deviations weighted equally) might reveal that *D* is the element at fault in the failure of connotative similarity to consistently determine generalization. The data on factors other than connotative meaning suggest, however, that Hypothesis *b* would probably also have to be modified in any instance.

Further research directed toward controlling or exploring the effect of context, meaningfulness, denotation, association, and other factors relevant to verbal generalization should reveal the extent to which this interpretation of the present data is adequate, and should make more precise the extent to which R_{SD} can be considered a valid measurement of "meaning" as treated by the Osgood theory.

Summary

A set of three hypotheses derived from Charles Osgood's approach to the nature and measurement of meaning were examined experimentally. The hypotheses concerned meaning as a representational mediation process, the validity of the Semantic Differential technique as a measurement of connotative meaning, and the Osgood-Suci index of profile similarity.

Semantic Differential profiles of words which had been previously scaled were used as the basis for determining connotative meaning. Six experimental studies were conducted to examine the hypothesis that mediated stimulus generalization will occur between stimuli with similar connotative meaning profiles. A lever-press response was taught to each member of clusters of connotatively similar words, and generalization was tested by presentation of additional words which were connotatively similar to the training words. The series of experiments included replication for stimulus items, alteration of verbal context of the stimuli, and exact replication of two experiments to determine the stability of generalized response. Supplementary analyses were done to determine the role of associative habits, usage frequency, meaningfulness, and denotative similarity.

The hypothesized generalization was obtained in each of 12 analyses of responses to groups of generalization stimuli. However, wide variability in generalized response was observed which could not be accounted for by the connotative meaning variable. A group of stimuli was identified for which generalized response unvaryingly failed to occur, in spite of the theoretical equivalence of these stimuli and the others.

The verbal context of the stimuli, connotative meaningfulness, and judged denotative meaning similarity were found to relate to the variation in generalized responses. Usage frequency and associative habits did not appear to relate to variability in generalization.

The results were interpreted as substantiating the mediation hypothesis and as lending some degree of construct validity to the Semantic Differential as a technique for measuring meaning. Alternative interpretations of the findings and implications for further research were discussed.

REFERENCES

BASTIAN, J. R. Response chaining in verbal transfer. Technical Report No. 13, 1957, University of Minnesota, Contract N8 onr-66216, Studies on the role of language in behavior.

BRAITHWAITE, R. *Scientific explanation.* Cambridge, England: Cambridge Univer., 1953.

COFER, C. N., JANIS, M. G., & ROWELL, M. M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32**, 266–269.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

DIVEN, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.

FOLEY, J. P., JR., & COFER, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 169–175.

HAAGEN, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Unpublished doctoral dissertation, State University of Iowa, 1943.

JENKINS, J. J., & RUSSELL, W. A. Research studies on the role of language in behavior. Technical Report, 1956, University of Minnesota, Contract N8 66216.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. A table of distances for the semantic atlas. *Amer. J. Psychol.*, 1959, **72**, 623–625.

KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, **67**, 37–96.

LUMSDAINE, A. A. Conditioned eyelid responses as mediating generalized conditioned finger reactions. *Psychol. Bull.*, 1939, **36**, 650. (Abstract)

MCCLELLAND, D. C., & HEATH, R. M. Retroactive inhibition as a function of degree of association between original and interpolated activities. *J. exp. Psychol.*, 1943, **33**, 420.

MCGEOCH, J. A., & MCDONALD, W. T. Meaningful relation and retroactive inhibition. *Amer. J. Psychol.*, 1931, **43**, 579–588.

MINK, W. D. Semantic generalization as related to word association. Unpublished doctoral dissertation, University of Minnesota, 1957.

MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592–603.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

NORMAN, W. T. Stability-characteristics of the semantic differential. *Amer. J. Psychol.*, 1959, **72**, 581-584.

O'CONNOR, J. *Born that way.* Baltimore: Williams & Wilkins, 1928.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 299–301.

OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.

OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325-338.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PAP, A. Reduction sentences and open concepts. *Methodos*, 1953, **5**, 3–30.

RAZRAN, G. Salivating, and thinking in different languages. *J. Psychol.*, 1935–36, **1**, 145–151.

RAZRAN, G. Semantic and phonetographic generalization of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, **39**, 642-652.

RIESS, B. F. Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26**, 238–240.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr-66216, Studies on the role of language in behavior.

RYAN, J. J. An experimental comparison of response transfer facilitated by meaningfully similar and associated verbal stimuli. Unpublished doctoral dissertation, University of Minnesota, 1957.

STORMS, L. H. Backward association in verbally mediated learning. Unpublished doctoral dissertation, University of Minnesota, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1944.

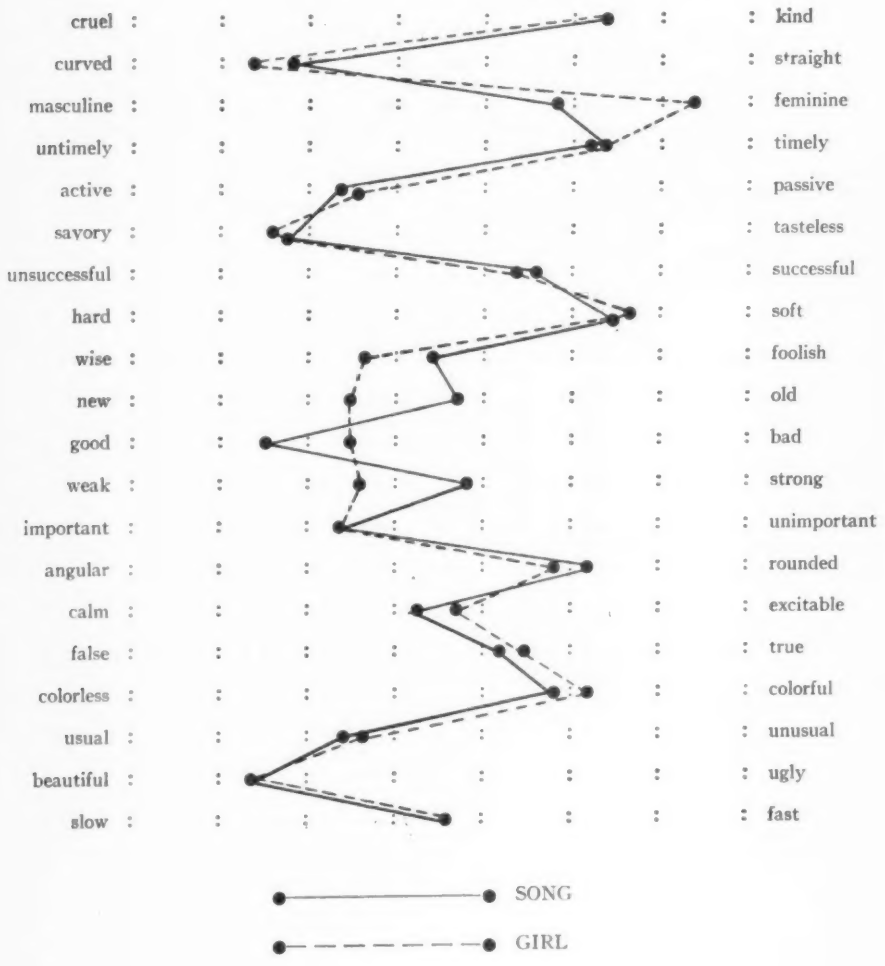
YOUNG, H. K., & UNDERWOOD, S. J. Transfer in verbal materials with dissimilar stimuli and response similarity varied. *J. exp. Psychol.*, 1954, **47**, 153–159.

YUM, K. S. An experimental test of the law of assimilation. *J. exp. Psychol.*, 1931, **14**, 68–82.

(Received April 23, 1960)

APPENDIX A

Semantic Differential Scales Used by Jenkins, Russell, and Suci


cruel : : : : : : : kind
curved : : : : : : : straight
masculine : : : : : : : feminine
untimely : : : : : : : timely
active : : : : : : : passive
savory : : : : : : : tasteless
unsuccessful : : : : : : : successful
hard : : : : : : : soft
wise : : : : : : : foolish
new : : : : : : : old
good : : : : : : : bad
weak : : : : : : : strong
important : : : : : : : unimportant
angular : : : : : : : rounded
calm : : : : : : : excitable
false : : : : : : : true
colorless : : : : : : : colorful
usual : : : : : : : unusual
beautiful : : : : : : : ugly
slow : : : : : : : fast
SONG
GIRL

APPENDIX B

FACTOR LOADINGS OF THE SCALES USED BY JENKINS, RUSSELL, AND SUCI

Scale	Factor								
	I	II	III	IV	V	VI	VII	VIII	h^2
	Evaluation	Potency	Activity	Stability	Tautness	Novelty	Receptivity	Unnamed	
Good–bad	1.00	.00	.00	.00	.00	.00	.00	.00	1.00
Timely–untimely	37	04	04	05	05	01	05	01	15
Kind–cruel	52	−28	00	16	−07	02	12	−07	41
Beautiful–ugly	52	−29	−02	03	−06	06	14	02	38
Successful–unsuccessful	51	08	29	06	00	06	09	12	38
Important–unimportant	38	04	31	04	00	−02	09	02	25
True–false	50	−03	01	29	−06	−01	00	05	34
Wise–foolish	57	06	11	22	−03	−02	10	05	40
Hard–soft	−24	97	00	00	00	00	00	00	1.00
Masculine–feminine	−14	47	03	−01	16	−05	−01	06	27
Strong–weak	30	40	10	12	00	−03	04	11	28
Active–passive	17	12	98	00	00	00	00	00	1.00
Excitable–calm	−15	03	26	−13	00	05	13	−04	13
Fast–slow	01	26	35	−05	15	−01	05	15	24
Angular–rounded	−12	26	16	−06	95	00	00	00	1.00
Straight–curved	08	12	14	06	27	05	−03	−02	12
New–old	20	−09	09	00	05	97	00	00	1.00
Unusual–usual	−04	02	03	00	03	25	12	03	08
Savory–tasteless	23	−12	18	04	−05	06	95	00	1.00
Colorful–colorless	20	−20	09	−04	−10	09	27	08	18

APPENDIX C

Sample Pages from Denotative Judgment Study

I

BODICE

Charming	Girl	Fragrant	Song
_____	_____	_____	_____

II

BODICE
LOVELY
PRETTY
WOMAN

Charming	Girl	Fragrant	Song
_____	_____	_____	_____